

The Fusion of Quadratic Detection Statistics Applied to Hyperspectral Imagery

David Stein¹, Alan Stocker², Scott Beaven³

¹SPAWAR Systems Center San Diego, Code D743, San Diego, CA 92152

²Space Computer Corporation, Santa Monica, CA 90404

³SPAWAR Systems Center San Diego, Code D855, San Diego, CA 92152

ABSTRACT

A variety of detection statistics have been developed and applied to hyperspectral imagery (HSI). The Reed Xiaoli (RX) algorithm is a generalized likelihood ratio test (GLRT) that uses local estimates of the spectral mean and spectral covariance. It satisfies an optimality criterion if, locally, the spectral data have a multivariate normal probability distribution. Alternatively, the stochastic expectation maximization (SEM) algorithm may be used to estimate the spectral mean values and spectral covariance matrices of a pre-determined number of classes. A detection statistic is computed by identifying each pixel with the class having maximal a posteriori probability and applying the GLRT detection statistic for that class. These algorithms are based on different models and provide different information about the imagery. For example, the RX algorithm seeks to identify local anomalies, and the SEM based detector attempts to discern those pixels that do not belong to one of the model classes. Thus we evaluate the improvement in detection performance that results from developing a joint RX-SEM decision criterion. The joint decision boundaries are obtained by modeling the output distribution of each of the algorithms and selecting a joint distribution that further incorporates the correlation between the RX and SEM detector output. The performance of the resulting fusion statistics are compared with the separate performance of the algorithms and AND/OR fusion rules.

1. INTRODUCTION

Fusion problems arise in surveillance if a scene is observed using multiple sensors, if data from a sensor is processed using various techniques, or if the scene is surveyed from diverse positions or at different times. Using multiple sensors is advantageous for detection and classification problems if the sensors, e.g., broadband imagers, foliage penetration radar, and spectral sensors covering various wavelengths, provide complimentary information. Furthermore, as each sensor may have better performance under certain conditions and as the characteristics of the background, target, and environment may be unknown, the best surveillance strategy for a given problem may be to fuse the outputs of several sensors and/or algorithms. Target detection algorithms are generally derived from models of the sensor data and an optimality criterion, such as the Neyman-Pearson rule, using approximations necessitated by limited knowledge of the background and target. Algorithms are often adapted to uncertain conditions by estimating parameters of the underlying model in-situ, however different modeling approaches may lead to fundamentally different techniques, and none of the methods may be universally superior. In these circumstances algorithm fusion may be used to resolve the inconsistencies of competing detection algorithms.

In this paper we develop transform methods to obtain joint-decision contours for a set of detection statistics, and we apply this technique to anomaly detection in hyperspectral imagery using the RX and SEM algorithms. The transforms are based on probability distributions of the output of the individual detection statistics, and their correlation. The performance of the transform approach to fusion is compared with AND and OR fusion rules and with a model selection approach. The performance of these algorithms is compared on three hyperspectral data sets.

2. RX AND SEM ALGORITHMS

The RX algorithm [7] is based on the assumptions that within a small region of a test pixel (i,j) the background has a multivariate normal density with mean μ_{ij} and spectral covariance matrix Γ_{ij} :

Report Documentation Page		
Report Date 00012001	Report Type N/A	Dates Covered (from... to) -
Title and Subtitle The Fusion of Quadratic Detection Statistics Applied to Hyperspectral Imagery	Contract Number	
	Grant Number	
	Program Element Number	
Author(s) Stein, David; Stocker, Alan; Beaven, Scott	Project Number	
	Task Number	
	Work Unit Number	
Performing Organization Name(s) and Address(es) SPAWAR Systems Center San Diego, Code D743, San Diego, CA 92152	Performing Organization Report Number	
Sponsoring/Monitoring Agency Name(s) and Address(es) Director, CECOM RDEC Night Vision and Electronic Sensors Directorate, Security Team 10221 Burbeck Road Ft. Belvoir, VA 22060-5806	Sponsor/Monitor's Acronym(s)	
	Sponsor/Monitor's Report Number(s)	
Distribution/Availability Statement Approved for public release, distribution unlimited		
Supplementary Notes The original document contains color images.		
Abstract		
Subject Terms		
Report Classification unclassified	Classification of this page unclassified	
Classification of Abstract unclassified	Limitation of Abstract UNLIMITED	
Number of Pages 9		

$$N(\mu_{ij}, \Gamma_{ij})(x_{ij}) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}(x_{ij} - \mu_{ij})^t \Gamma_{ij}^{-1} (x_{ij} - \mu_{ij})\right). \quad (1)$$

The RX detection statistic [7] is derived as a generalized likelihood ratio test:

$$\Lambda(y) = \frac{\max_{\omega \in \Omega_1} p_1(y, \omega)}{\max_{\omega \in \Omega_0} p_0(y, \omega)}, \quad (2)$$

where $p_i(\cdot, \omega)$ is the probability density function of the observation under hypothesis i

(H_0 : target absent; H_1 : target present) given the parameter ω . In application to RX, $\omega = [b, \Gamma]$

where b is the signature of the target spectrum, and Γ is the spectral covariance matrix. For a single pixel target the RX statistic may be represented as

$$RX(y) = (y - \mu)^t \Gamma^{-1} (y - \mu). \quad (3)$$

The RX statistic is defined more generally for spatially extended targets by [7]

$$R(\bar{x}_{ij}) = \sum_{\alpha, \beta} w_{\alpha\beta} (\bar{x}_{i-\alpha, j-\beta} - \bar{\mu}_{i-\alpha, j-\beta})^t \Gamma_{i-\alpha, j-\beta}^{-1} (\bar{x}_{i-\alpha, j-\beta} - \bar{\mu}_{i-\alpha, j-\beta}). \quad (4)$$

The stochastic expectation and maximization algorithm (SEM) [1,4] has been applied to segment hyperspectral imagery into sets of classes each of which is described by a multivariate Gaussian pdf. Thus the hyperspectral image is modeled as having a Gaussian mixture pdf

$$\bar{x} \sim \sum_{j=1}^m \rho_j N(\mu_j, \Gamma_j) \text{ such that } 0 \leq \rho_j \text{ and } \sum_{j=1}^m \rho_j = 1. \quad (5)$$

A detection statistic has been constructed on the basis of SEM segmentation. Pixel (i, j) is assigned to the SEM class $\psi_{ij} = k$ such that the conditional probability that observation x_{ij} came from class k is maximal i.e.

$$k = \arg(\max_h p(\psi_{ij} = h | x_{ij}, \{\rho_j, \mu_j, \Gamma_{js} | 1 \leq j \leq m\}) \quad (6)$$

The SEM detection statistic operating on x_{ij} is then the RX algorithm for the given class, i.e. it is the class conditional GLRT

$$S(\bar{x}_{ij}) = (\bar{x}_{ij} - \mu_k)^t \Gamma_k^{-1} (\bar{x}_{ij} - \mu_k) \quad (7).$$

Gamma mixture distributions were used to model the output of the RX and SEM algorithms. The gamma density with shape parameter ν and scale parameter a is given by

$$F_{\nu, a}(x) = \frac{x^{\nu-1} e^{-x/a}}{a^\nu \Gamma(\nu)}. \quad (8)$$

A gamma mixture density has the form

$$F(x) = \sum_{k=1}^m \rho_k F_{\nu_k, a_k}(x), \text{ where } \sum_{k=1}^m \rho_k = 1 \text{ and } \rho_k \geq 0. \quad (9)$$

At the pixel level the RX, Eq. 3, and SEM, Eq. 7, algorithms apply a quadratic form

$$Q(\mu, \Gamma)(\bar{x}) = (\bar{x} - \bar{\mu})^t \Gamma^{-1} (\bar{x} - \bar{\mu}) \quad (10)$$

to each observation. Thus, if the background is normally distributed and if parameters are known with sufficient accuracy, the output of RX applied to background only data would have a chi-squared distribution on n degrees of freedom (DOF) where n is the number of linearly independent dimensions of the data [8]. The chi-squared distribution on n DOF is the gamma distribution with $\nu = n/2$, and $a = 1$. If the background regions have Gaussian mixture distributions or if there is significant parameter estimation error, then the RX output will have a gamma mixture distribution [6]. Similarly, if the SEM algorithm perfectly assigned pixels to Gaussian classes, the output of SEM from each class would be a chi-squared distribution. Classification and parameter estimation errors, however, can lead to SEM having a gamma mixture distribution. The background may not be well fit by a Gaussian mixture distribution, in which case the Gamma mixture distribution may not provide the best approach to modeling the detector outputs. The fusion method defined below is not dependent on any particular form of the marginal densities and can be implemented using non-parametric descriptions of the pdf. However, the gamma mixture distributions, with $m=2$, provide good descriptions of the data utilized in this paper. For example, Figure 1 compares the empirical cumulative distribution of the SEM algorithm applied to desert VNIR data and the estimated 2-state gamma mixture distribution.

As indicated above, we anticipate that the distribution of the RX algorithm will depend in part on whether or not the neighborhood of a test pixel is well modeled by a Gaussian distribution or a Gaussian mixture distribution. We use the BHEP test [2] to evaluate the goodness of fit of the multivariate normal distribution to the local background. The BHEP test statistic compares the empirical characteristic function of the data, transformed to zero mean and identity covariance, with the characteristic function of the zero-mean identity-covariance normal distribution. Let the sample data to be tested be $\{X_1, \dots, X_n\} \subset R^d$, and let $Y_i = H^{-1}(X_i - \mu)$, where $\mu = E(X)$ and $\Gamma = HH^t = \text{cov}(X)$. The empirical characteristic function of Y is

$$\psi_n(t) = \frac{1}{n} \sum_{j=1}^n \exp(itY_j), \quad (11)$$

and the BHEP test statistic is

$$D_{n,\beta} = \int_{R^d} \left| \psi_n(t) - \exp\left(-\frac{1}{2}\|t\|^2\right) \right|^2 \varphi_\beta(t) dt, \quad (12)$$

where

$$\varphi_\beta(t) = (2\pi\beta^2)^{-d/2} \exp\left(-\frac{\|t\|^2}{2\beta^2}\right).$$

The test is consistent, invariant under affine transformations, and applicable to any number of samples and data dimensions. The first three moments of the limiting distribution, as the number of samples approaches infinity, are known and can be used to approximate thresholds of the test statistic corresponding to prescribed probabilities of type I error. The application of this test also forms the basis for model-selection based fusion described below.

3. FUSION RULES

Model selection and joint-decision fusion approaches are developed and applied to RX-SEM output. A fusion rule partitions the RX-SEM space into regions R_0 and R_1 such that if $(r(x), s(x)) \in R_0$ then observation x is declared to be target free, and if $(r(x), s(x)) \in R_1$ then observation x is declared to come from a target. Since RX and SEM are positive valued, these regions may be defined by a mapping $\delta : R^{+2} \rightarrow \{0,1\}$ such that $R_i = \delta^{-1}(i)$, where $R^+ = [0, \infty)$.

The model selection rule is defined as follows. For each observation x_{ij} , let $b = b(x_{ij}) = BHEP(N(i, j))$ be the BHEP test, Equation 12, applied to the reference data in a neighborhood $N(i, j)$ of pixel (i, j) , and let τ_{BHEP}, τ_{RX} , and τ_{SEM} be thresholds for BHEP, RX, and SEM respectively. Then

$$\delta_{MS}(r, s, b) = \begin{cases} 1 & \text{if } (b \leq \tau_{BHEP} \text{ and } r > \tau_{RX}) \text{ or } (b > \tau_{BHEP} \text{ and } s > \tau_{SEM}) \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

The AND and OR fusion rules are based on the marginal distributions of RX and SEM output. For each probability of false alarm, α , let $\tau_{R(\alpha)}$ and $\tau_{S(\alpha)}$ be the corresponding thresholds of RX and SEM, respectively. Then

$$\delta_{AND}(r, s) = \begin{cases} 0, & \text{if } r \leq \tau_r(\alpha) \text{ or } s \leq \tau_s(\alpha) \\ 1, & \text{otherwise} \end{cases} \quad (14)$$

$$\delta_{OR}(r, s) = \begin{cases} 0, & \text{if } r \leq \tau_r(\alpha) \text{ and } s \leq \tau_s(\alpha) \\ 1, & \text{otherwise} \end{cases} \quad (15)$$

The joint-density method utilizes a joint pdf defined on (R, S) . Let $f(r, s)$ be a joint RX-SEM density, then f can be used to define decision regions, R_0 and R_1 , as follows. For each $c \in [0, \infty)$ define

$$\delta_{JD}(r, s) = \begin{cases} 0, & \text{if } f(r, s) \geq c \\ 1, & \text{otherwise} \end{cases} \quad (16)$$

Note that δ_{JD} is a likelihood ratio in case the distribution of RX-SEM under the target present hypothesis is uniform on a bounded region R such that $(r, s) \in R^c \Rightarrow \delta(r, s) = 1$, where R^c denotes the complement of R .

In the present work, the joint densities have been constructed using a transform approach. Let $x = (x_1, \dots, x_k) \in R^k$ be a random vector such that the cumulative distribution function (CDF) of x_k is Ψ_k . Let $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp(-\frac{1}{2}t^2)dt$ be the zero-mean, unit-variance normal CDF. Then $y_k = \Psi_k(x_k)$ has a uniform distribution, and $z_k = \Phi^{-1}(y_k)$ has a zero-mean, unit-variance normal distribution. Let $C = \text{cov}(z_1, \dots, z_k)$, and

$$p(z) = \frac{1}{(2\pi)^{d/2} \|C\|^{1/2}} \exp(-\frac{1}{2} z^t C^{-1} z). \quad (17)$$

Define

$$f(x) = p(T(x)) \|\nabla T(x)\|, \quad (18)$$

where $T(x) = (\Phi^{-1}(\Psi_1(x_1)), \dots, \Phi^{-1}(\Psi_k(x_k)))$ and $\|\nabla T(x)\|$ is the determinant of the Jacobian of T . Then f is a probability density function on R^k such that the distribution of x_k induced by f , the k^{th} marginal distribution, is Ψ_k and C , the normal score correlation of f , is an asymptotically unbiased estimator of the normal score correlation of the underlying distribution [3]. Furthermore, f is the maximum entropy density among pdfs having prescribed marginals and normal score correlation [3]. f is the pullback of p via T .

A modification of this approach can be utilized to obtain a standard set of marginal densities. Let $(\Lambda_1, \dots, \Lambda_k)$ be a preferred set of k marginal distributions, and define $w_j = \Lambda_j^{-1}(\Psi_j(x_j))$. Then the j^{th} marginal distribution of $w = (w_1, \dots, w_k)$ is Λ_j . Define pdf g to be the pullback of p via $S(w) = \Phi^{-1}(\Lambda_1(w_1)), \dots, \Phi^{-1}(\Lambda_k(w_k))$. For example, Moran [5] uses this construction for $k = 2$ and Λ_j a gamma distribution. We are evaluating the relative advantages of standardizing the marginal distributions to exponential form.

4. APPLICATIONS

These techniques have been applied to three hyperspectral data sets and the results are displayed in Figures 2-12. In these figures FR refers to fusion using the joint-density approach in which separate densities are estimated for the sets B_0 , and B_1 defined by $B_0 = \{x \mid b(x) \leq \tau_{BHEP}\}$ and $B_1 = \{x \mid b(x) > \tau_{BHEP}\}$. FG refers to fusion using the joint-density approach in which a global density is fit to (R, S) output. MS refers to fusion using the model selection approach. The distributions of RX and SEM output were fit to 2-term gamma mixture densities using the expectation-maximization algorithm. The targets are indicated as red diamonds

Data set one is VNIR hyperspectral data collected over a desert. Figure 2 is a scatter plot of spatially associated RX-SEM output with an overlay of the contours of the probability density obtained using the transform method defined above. Figure 3 compares ROC curves of SEM, RX and FR. The six outermost targets are detected without false alarms by SEM and FR, whereas RX incurs a significant number of false alarms to detect these targets. Figure 4 compares the performance obtained using fusion rules FG, FR, and MS. FR is able to detect six targets without false alarms, whereas FG can only detect five without false alarm. Furthermore, FR incurs approximately an order of magnitude fewer false alarms than MS to detect the seventh target. Figure 5 compares the performance of fusion rules OR, AND, and FR. OR and FR are comparable and detect the outermost targets without false alarms, whereas AND incurs a significant number of false alarms at threshold settings that detect these six targets.

Data set two, which is VNIR hyperspectral data collected over a forest, is similarly analyzed in Figures 6-9. The six (PD=0.33) outermost targets in Figure 6 are detected with fewer false alarms using one of the fusion approaches than using either RX or SEM alone as evidenced in the ROC curves. At this level, from Figure 8, one sees that FR has approximately half as many false alarms as FG and MS. From Figure 9, at PD=0.3, one sees that FR has about an equal number of false alarms as AND and approximately half as many as OR. The next three targets are substantially further down in the clutter in Figure 6, and the number of false alarms incurred to detect them goes up by approximately two orders of magnitude using any of the fusion techniques. From this point to detection of all targets the fusion result is comparable to the better of the two algorithms, SEM, but neither is satisfactorily separating the targets from the clutter.

Data set three, which is LWIR hyperspectral data from a forest, is analyzed in Figures 10-12. The data set contains seven targets, and several of the targets are represented by more than one diamond in Figure 10. We see, from Figure 11, significant fusion gain, a reduction in the number of false alarms by 1.5 and 2 orders of magnitude, using FR rather than RX or SEM, respectively, at threshold levels sufficient to detect the four outermost targets. The other three targets are well inside the clutter, and the performance of FR is comparable to RX, which in this case outperforms SEM. From Figure 12 we see that FR outperforms OR and AND in detecting the four outermost targets by an order of magnitude in the number of false alarms, and their performance is comparable at higher probabilities of false alarm. The model selection fusion approach would default to the SEM algorithm in this case as all pixels evaluated failed the BHEP test.

5. CONCLUSIONS/FUTURE DIRECTIONS

This study has demonstrated improved performance by following RX and SEM processing with a fusion algorithm. At thresholds such that either RX or SEM has fewer than 10-100 false alarms per km^2 fusion has been shown to reduce false alarms by 0.25 to 2 orders of magnitude. Furthermore, if thresholds are set so that both algorithms produce more than approximately 100 false alarms per km^2 , the performance of the fusion algorithm was comparable to or better than either RX or SEM alone.

This work has also shown that the FR fusion algorithm produces more consistent results than either of the other algorithms considered in this study. Table 1 shows the best algorithm from the set of choices for each data set. If the choices are SEM and RX, then an algorithm selection criterion should be developed to determine the conditions under which each algorithm should be used, as SEM is the algorithm of choice for data sets one and two, while RX achieves better results for data set 3. The model selection criterion based on the BHEP test worked quite well for data sets 1 and 2, as evidenced in Figures 4 and 8. However, this test would have selected SEM for all of data set 3, and RX was the better algorithm to use for these data. Similarly AND is preferred over OR on data set 2, but OR is preferred over AND on data sets 1 and 3. Thus if these fusion rules are to be adopted, then a selection criterion needs to be developed. When the choices include FR it is the preferred approach, and its performance is very similar to OR on data set 1 and to AND on data set 2. On data set 3, FR and FG are identical as all pixels lie in the non-Gaussian class. On data sets one and two, however, FR is preferred over FG.

Table 1. The optimal algorithm from the set of choices for each data set.

Algorithm Choices	Data Set 1	Data Set 2	Data Set 3
SEM, RX	SEM	SEM	RX
SEM, RX, AND, OR	OR	AND	OR
SEM, RX, AND, OR, MS, FR, FG	OR~FR	AND~FR	FR=FG

There are many issues to address in algorithm and sensor fusion. We are investigating alternate means of modeling the marginal distributions and constructing the joint density. We are evaluating the relative merits of pixel level and spatially associated fusion. We are analyzing the computational efficiency and accuracy of various parameter estimation techniques. We will be incorporating other detection algorithms, and we intend to apply these techniques to certain sensor fusion problems.

REFERENCES

- [1] S. G. Beaven and L. E. Hoff, "Comparison of SEM and Linear Unmixing Approaches for Classification of Spectral Data," *Proceedings of SPIE*, vol. 3753, Imaging spectrometry V, M. R. Descour and S. S. Shen (eds.), 19-21 July 1999, Denver, CO, pp. 300—307.
- [2] N. Henze and T. Wagner, "A New Approach to the BHEP Tests for Multivariate Normality," *Journal of Multivariate Analysis*, no. 62, 1997, pp. 1—23.
- [3] M. J. W. Jansen, "Maximum Entropy Distributions with Prescribed Marginals and Normal Score Correlations," in *Distributions with given Marginals and Moment Problems*, V. Benes and J. Stepan (eds.) Kluwer Academic Publishers, 1997, pp. 87—92.
- [4] P. Masson and W. Pieczynski, "SEM Algorithm and Unsupervised Statistical Segmentation of Satellite Images," *IEEE Transaction on Geoscience and Remote Sensing*, Vol. 31, No. 3, May 1993.
- [5] P. A. P. Moran, "Statistical inference with bivariate gamma distributions," *Biometrika*, Vol 56, No. 3, 1969, pp. 627—634.
- [6] S. J. Press, "Linear combinations of Non-Central Chi-Square Variates," *Annals of Math Stat.* No. 37, 1966, pp. 480—487.
- [7] I. S. Reed and X. Yu, "Adaptive Multiple-Band CFAR Detection of an Optical Pattern with Unknown Spectral Distribution," *IEEE Transactions on Acoustics, speech, and Signal Processing*, Vol 38 no 10, October 1990, pp. 1760—1770.
- [8] A. D. Whalen, *Detection of Signals in Noise*, Academic Press, New York, 1971, Chapter 4.

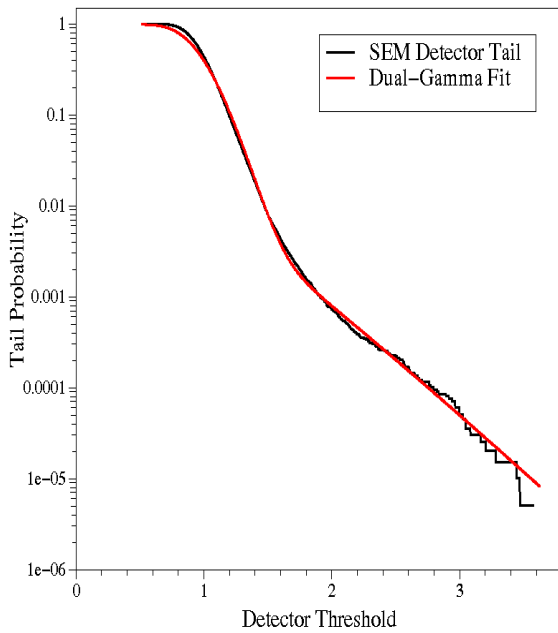


Figure 1. Two-term gamma mixture density fit to SEM detector output from desert background data having parameters $\rho_1 = 0.962$, $\rho_2 = 0.038$, $v_1=15.7$, $v_2=2.44$, $a_1=0.035$, $a_2=0.305$.

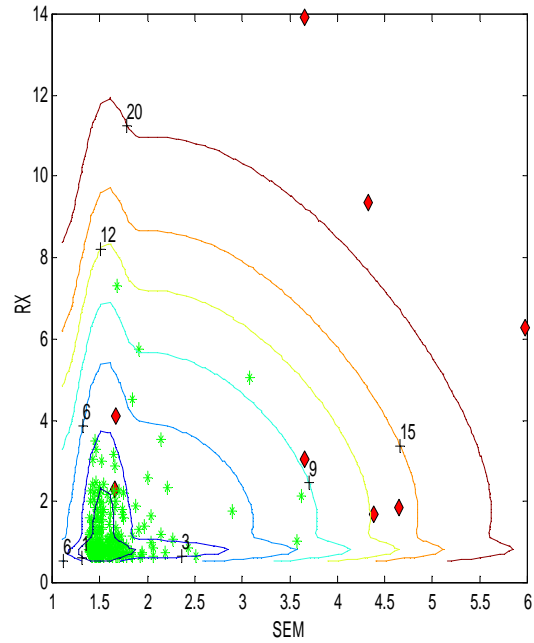


Figure 2. Scatter plot of spatially associated RX-SEM output from desert data and contours of minus the natural logarithm of the joint density.

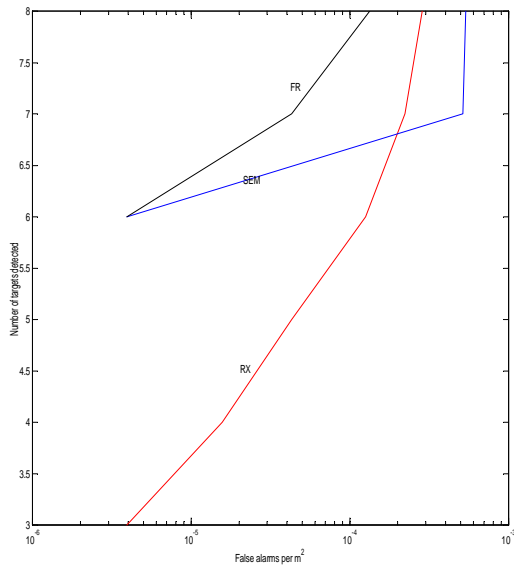


Figure 3. ROC comparison of RX, SEM, and FR applied to desert VNIR data.

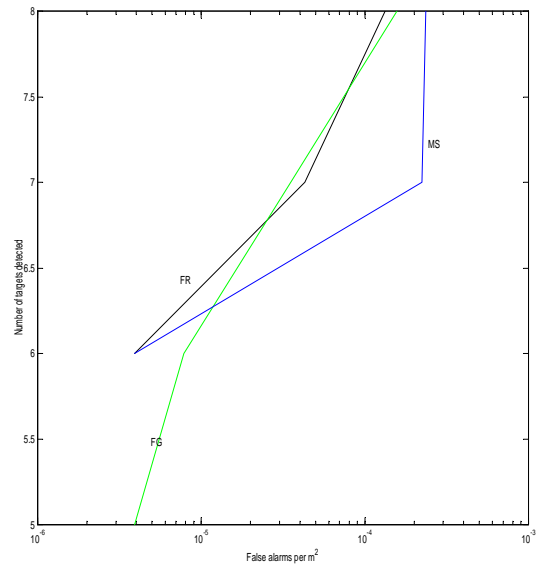


Figure 4. ROC comparison of fusion methods FR, FG, and MS applied to desert VNIR data.

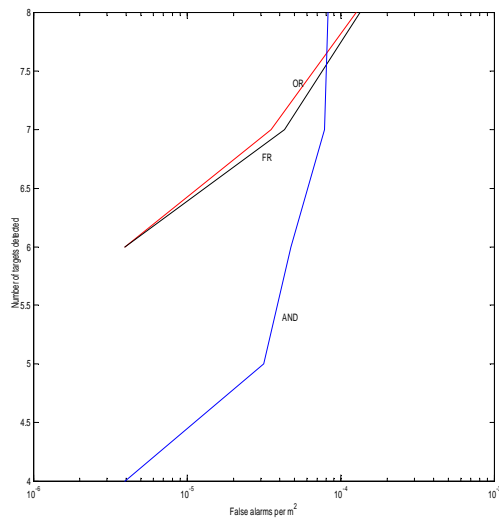


Figure 5. ROC comparison of fusion methods FR, AND, and OR applied to desert VNIR data.

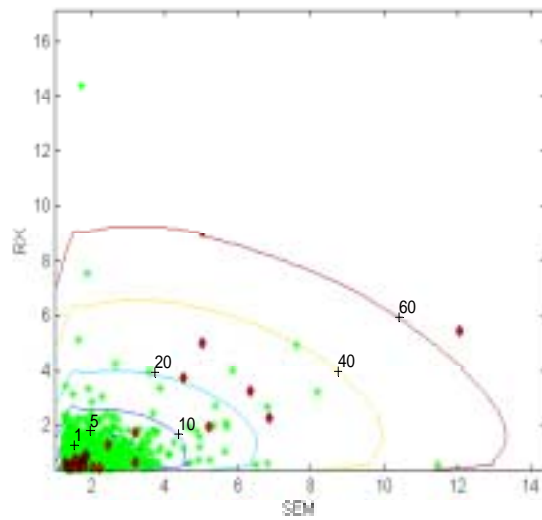


Figure 6. Scatter plot of spatially associated RX-SEM output from VNIR forest data and contours of minus the natural logarithm of the joint density.

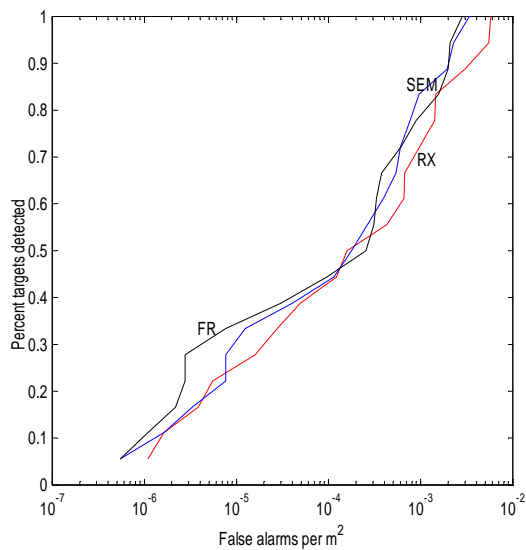


Figure 7. ROC comparison of FR, RX, and SEM applied to forest VNIR data.

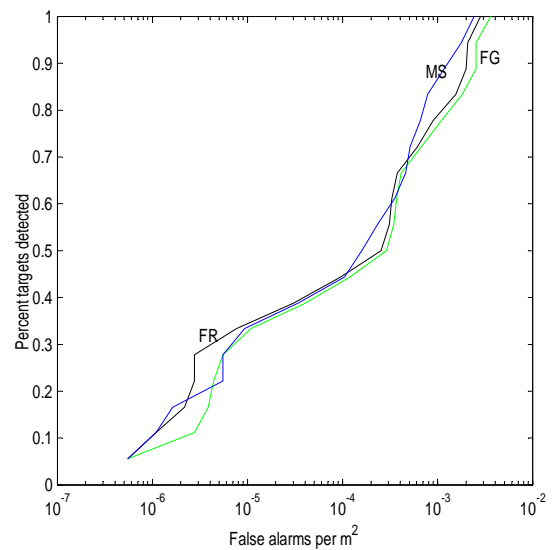


Figure 8. ROC comparison of fusion methods FR, MS, and FG applied to forest VNIR data.

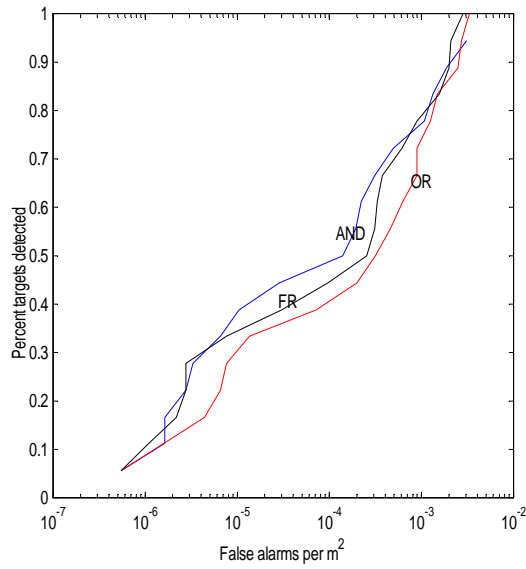


Figure 9. ROC comparison of fusion methods FR, AND, and OR applied to forest VNIR data.

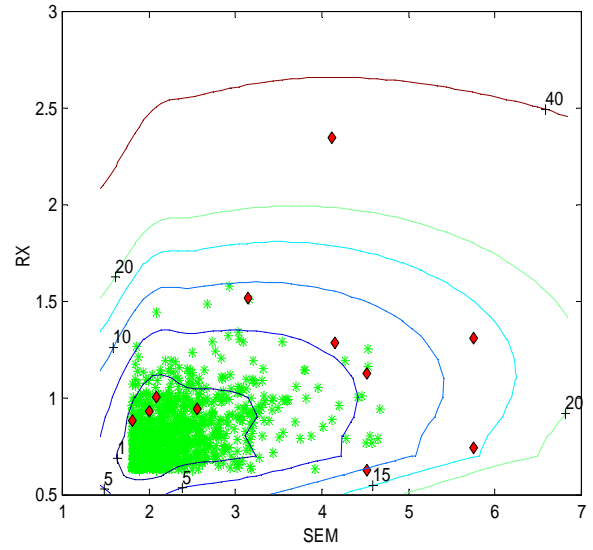


Figure 10. Scatter plot of spatially associated RX-SEM output from LWIR forest data and contours of minus the natural logarithm of the joint density.

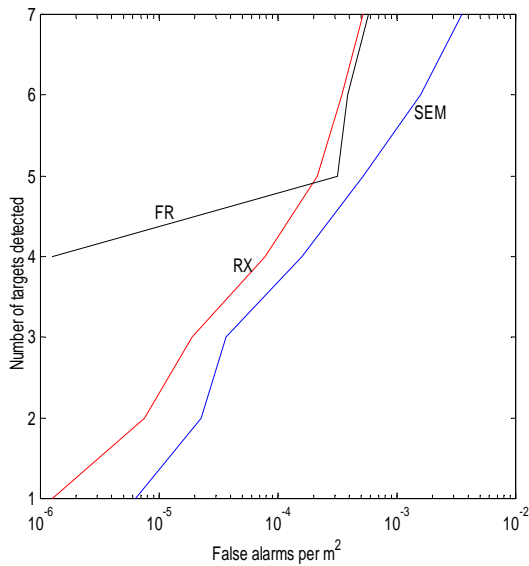


Figure 11. ROC comparison of FR, RX, and SEM applied to forest LWIR data.

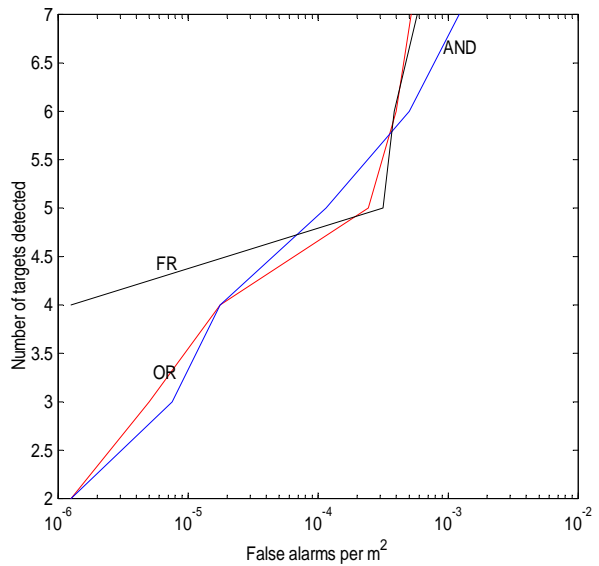


Figure 12. ROC comparison of fusion methods FR, AND, and OR applied to forest LWIR data.